



Symplectic/Contact Geometry Related to Bayesian Statistics [†]

Atsuhide Mori

Department of Mathematics, Osaka Dental University, Osaka 573-1121, Japan; mori-a@cc.osaka-dent.ac.jp

[†] Presented at the 5th International Electronic Conference on Entropy and Its Applications, 18–30 November 2019; Available online: <https://ecea-5.sciforum.net/>.

Published: 17 November 2019



Abstract: In the previous work, the author gave the following symplectic/contact geometric description of the Bayesian inference of normal means: The space \mathbb{H} of normal distributions is an upper halfplane which admits two operations, namely, the convolution product and the normalized pointwise product of two probability density functions. There is a diffeomorphism F of \mathbb{H} that interchanges these operations as well as sends any e-geodesic to an e-geodesic. The product of two copies of \mathbb{H} carries positive and negative symplectic structures and a bi-contact hypersurface N . The graph of F is Lagrangian with respect to the negative symplectic structure. It is contained in the bi-contact hypersurface N . Further, it is preserved under a bi-contact Hamiltonian flow with respect to a single function. Then the restriction of the flow to the graph of F presents the inference of means. The author showed that this also works for the Student t -inference of smoothly moving means and enables us to consider the smoothness of data smoothing. In this presentation, the space of multivariate normal distributions is foliated by means of the Cholesky decomposition of the covariance matrix. This provides a pair of regular Poisson structures, and generalizes the above symplectic/contact description to the multivariate case. The most of the ideas presented here have been described at length in a later article of the author.

Keywords: information geometry; Poisson structure; symplectic structure; contact structure; foliation; Cholesky decomposition

1. Introduction

We work in the C^∞ -smooth category. A manifold U embedded in the space of probability distributions inherits a separating premetric $D : U \times U \rightarrow \mathbb{R}_{\geq 0}$ from the relative entropy, which is called the Kullback–Leibler divergence. The geometry of (U, D) is studied in the information theory. The information geometry [1] concerns the infinitesimal behavior of D . In the case where U is the space of univariate normal distributions, we regard U as the half plane $\mathbb{H} = \mathbb{R} \times \mathbb{R}_{>0} \ni (m, s)$, where m denotes the mean and s the standard deviation. Since the convolution of two normal densities is a normal density, it induces a product $*$ on \mathbb{H} , which we call the convolution product. On the other hand, since the pointwise product of two normal densities is proportional to a normal density, it induces another product \cdot on \mathbb{H} , which we call the Bayesian product. The first half of this presentation is devoted to the geometric description of Bayesian statistics including this product.

On the other hand, the current statistics lies not only in probability theory, but also in information theory. The author [2] found a symplectic description of the statistics of univariate normal distributions which is simultaneously based on these theories. Precisely, on the product $\mathbb{H} \times \mathbb{H}$ with coordinates

(m, s, M, S) , we take the positive and negative symplectic forms $d\lambda_{\pm}$ with the fixed primitives $\lambda_{\pm} = \frac{dm}{s} \pm \frac{dM}{S}$. Then the Lagrangian surfaces

$$F_{\varepsilon} = \left\{ (m, s, M, S) \in \mathbb{H} \times \mathbb{H} \mid \frac{m}{s} + \frac{M - \varepsilon}{S} = 0, \quad sS = 1 \right\} \quad (\varepsilon \in \mathbb{R})$$

with respect to $d\lambda_{-}$ foliate the hypersurface $N = \{sS = 1\}$. For each $\varepsilon \in \mathbb{R}$, the leaf F_{ε} is the graph of a diffeomorphism of \mathbb{H} which sends any geodesic to a geodesic with respect to the e-connection. Further, in the case where $\varepsilon = 0$, the diffeomorphism interchanges the products $*$ and \cdot , namely,

$$\begin{cases} (m, s, M, S) \in F_0 \\ (m', s', M', S') \in F_0 \end{cases} \Rightarrow \begin{cases} ((m, s) * (m', s'), (M, S) \cdot (M', S')) \in F_0 \\ ((m, s) \cdot (m', s'), (M, S) * (M', S')) \in F_0. \end{cases}$$

Thus, the iteration of $*$ in the first factor of $\mathbb{H} \times \mathbb{H}$ corresponds to that of \cdot in the second factor. The primitives λ_{\pm} , the hypersurface N , the foliation $\{F_{\varepsilon}\}_{\varepsilon \in \mathbb{R}}$ and the leaf F_0 are preserved under the diffeomorphism $\varphi_{\zeta} : (m, s, M, S) \mapsto (\zeta m, \zeta s, \zeta^{-1}M, \zeta^{-1}S)$ for any $\zeta \in \mathbb{R}_{>0}$. This map appears in the construction of Hilbert modular cusps by Hirzebruch [3]. Further the function $f : \mathbb{H} \times \mathbb{H} \rightarrow \mathbb{R}_{\geq 0}$ which is defined by $f(m, s, M, S) = D(m, s, m', s')$ for $(m', s', M, S) \in F_0$ is also preserved under φ_{ζ} . On the other hand, the hypersurface N inherits the mutually transverse pair of the contact structures $\ker(\lambda_{\pm}|_N)$, which we call the bi-contact structure. Let X be the contact Hamiltonian vector field of $\lambda_{+}|_N$ with respect to the function $\frac{m}{s}$, i.e., the unique contact vector field satisfying $\lambda_{+}|_N(X) = \frac{m}{s}$. Then X is also the contact Hamiltonian vector field of $\lambda_{-}|_N$ with respect to the same function $\frac{m}{s}$. We call such a vector field a bi-contact Hamiltonian vector field. There is a non-trivial bi-contact Hamiltonian vector field on (N, λ_{\pm}) which is tangent to the leaf F_0 . It is the one for the above function $\frac{m}{s}$ up to constant multiple. We may regard its restriction to F_0 as a vector field on the second factor of $\mathbb{H} \times \mathbb{H}$ since F_0 is the graph of a diffeomorphism. Surprisingly, this vector field is tangent to a foliation by e-geodesics, and each leaf is closed under the Bayesian product. Further, the author [4] showed that a similar vector field on the squared space of Student's t -distributions can provide an indication of "geometric smoothness" in actual data smoothing.

In the second half of this presentation, we generalize the above description to the multivariate case. It is straightforward except that we use the Cholesky decomposition to foliate the squared space of n -variate normal distributions. Here the leaves are $4n$ -dimensional submanifolds carrying two symplectic structures. They form a pair of Poisson structures on the squared space.

This short paper provides the calculation results and sketches the mathematical ideas. For the precise descriptions and the proofs, see the article [5] by the author.

2. Results

2.1. Bayesian Information Geometry

In this subsection, we generalize the setting of the information geometry. Take a smooth family of volume forms with finite total volumes on \mathbb{R}^n . We regard each of the volume forms as a point of a manifold \mathcal{M} , namely, a point $y \in \mathcal{M}$ presents a volume form $\rho_y d\text{Vol}$ smoothly depending on y . Let \mathcal{V} be the space of volume forms with finite total volumes on \mathcal{M} . We take a volume form V in \mathcal{V} . Given a point z on \mathbb{R}^n , we regard the value $\rho_y(z)$ of the density as a function $\rho(z) : y \mapsto \rho_y(z)$, and multiply the volume form V by the function $\rho(z)$. This defines the updating map

$$\varphi : \mathbb{R}^n \times \mathcal{V} \ni (z, V) \mapsto \rho(z)V \in \mathcal{V}. \quad (1)$$

We notice that a volume form with finite total volume is proportional to a probability measure. Thus the function $\rho(z)$ is proportional to the likelihood, and System (1) presents Bayes' rule.

A proper subset $\tilde{\mathcal{U}} \subset \mathcal{V}$ is called a (generalized) conjugate prior if it satisfies

$$\varphi(\mathbb{R}^n \times \tilde{\mathcal{U}}) \subset \tilde{\mathcal{U}}. \quad (2)$$

Suppose that we have a conjugate prior $\tilde{\mathcal{U}}$ which is a smooth manifold, and further that, by using the hypersurface $\mathcal{U} = \{V \in \tilde{\mathcal{U}} \mid \int_{\mathbb{R}^n} V = 1\}$, it can be written as $\tilde{\mathcal{U}} = \{kV \mid V \in \mathcal{U}, k > 0\}$. We define on $\tilde{\mathcal{U}}$ the following “distance” \tilde{D} , which satisfies non of the axioms of distance.

$$\tilde{D}(V_1, V_2) = \int_{\mathbb{R}^n} V_1 \ln \frac{V_2}{V_1} \quad (\text{the relative entropy}) \quad (3)$$

Note that the restriction $\tilde{D}|_{\mathcal{U} \times \mathcal{U}} = D$ satisfies the separation axiom, and is called the Kullback–Leibler divergence. We have to fix the coordinate in the k -direction which presents the time. Then we write the quadratic term of the Taylor expansion of $\tilde{D}(P, P + dP) + \tilde{D}(P + dP, P)$ as $\sum_{i,j} \tilde{g}_{ij} dP^i dP^j$, where $\tilde{g}_{ij} = \tilde{g}_{ji}$. Suppose that $\tilde{g} = [\tilde{g}_{ij}]$ is a metric on $\tilde{\mathcal{U}}$. Let $\tilde{\nabla}^0$ be the Levi-Civita connection with respect to \tilde{g} . We write the cubic term of the expansion of $3\tilde{D}(P, P + dP) - 3\tilde{D}(P + dP, P)$ symmetrically as $\sum_{i,j,k} \tilde{T}_{ijk} dP^i dP^j dP^k$. This defines the line of (generalized) α -connections $\tilde{\nabla}^\alpha = \tilde{\nabla}^0 - \alpha \tilde{g}^* \tilde{T}$ with affine parameter $\alpha \in \mathbb{R}$, where $\tilde{g}^* T$ denotes the contraction $\sum_l \tilde{g}^{kl} \tilde{T}_{ijl}$ by the contravariant metric $\tilde{g}^{-1} = [\tilde{g}^{ij}]$. Note that $\tilde{\nabla}^\alpha$ has no torsion. Restricting all of the above notions with tilde to the hypersurface $\mathcal{U} \subset \tilde{\mathcal{U}}$, we obtain the notions without tilde in the usual information geometry [1]. Here \mathcal{U} can be identified with a space U of probability distributions.

2.2. The Geometry of Normal Distributions

In this subsection we consider the space U of multivariate normal distributions. The pair of a vector $\mu = (\mu_i)_{1 \leq i \leq n} \in \mathbb{R}^n$ and an upper triangular matrix $C = [c_{ij}]_{1 \leq i, j \leq n} \in \text{Mat}(n, \mathbb{R})$ with positive diagonal entries determines an n -variate normal distribution by declaring that μ presents the mean and $C^T C$ the Cholesky decomposition of the covariance matrix. We put

$$\sigma_i = c_{ii} \quad \text{and} \quad r_{ij} = \frac{c_{ij}}{c_{ii}} \quad (i, j \in \{1, \dots, n\}), \quad \text{i.e.,} \quad C = \text{diag}(\sigma) [r_{ij}].$$

Note that $[r_{ij}]$ is unitriangular, i.e., it is a triangular matrix whose diagonal entries are all 1. Considering $\sigma \in \mathbb{R}^n$ and $r = (r_{ij})_{1 \leq i < j \leq n} \in \mathbb{R}^{n(n-1)/2}$ as parameters, we can write the probability density of the n -variate normal distribution at $P = (\mu, \sigma, r) \in U = \mathbb{R}^n \times (\mathbb{R}_{>0})^n \times \mathbb{R}^{n(n-1)/2}$ as

$$p(x) = \frac{1}{\sqrt{(2\pi)^n |\sigma|}} \exp \left(-\frac{1}{2} \|C(\sigma, r)^{-T} (x - \mu)\|^2 \right) \quad (x \in \mathbb{R}^n).$$

Then the relative entropy defines the premetric

$$D(P, Q = (\mu', \sigma', r')) = \frac{\|C(\sigma', r')^{-T} (\mu' - \mu)\|^2}{2} + \frac{\|C(\sigma, r)C(\sigma', r')^{-1}\|^2 - n}{2} - \sum_{i=1}^n \ln \frac{\sigma_i}{\sigma'_i},$$

where $\|\cdot\|^2$ denotes the sum of squares (i.e., $\|\cdot\|$ the Frobenius norm). Thus,

$$D(P + \Delta P, P) = \frac{\|C^{-T} \Delta \mu\|^2}{2} + \frac{\|\Delta C C^{-1}\|^2}{2} + \text{tr}(\Delta C C^{-1}) - \ln |1_n + \Delta C C^{-1}|,$$

where 1_n is the unit, and ΔC the difference $C(\sigma + \Delta\sigma, r + \Delta r) - C(\sigma, r)$. Let r^{ij} be the entries of the inverse matrix of $[r_{ij}]$. Then we have

$$(\text{the } ij\text{-entry of } \Delta C C^{-1}) = \begin{cases} \frac{\Delta\sigma_i}{\sigma_i} & (i = j) \\ \frac{\sigma_i + \Delta\sigma_i}{\sigma_j} \sum_{k=i+1}^j r^{kj} \Delta r_{ik} & (i < j) \\ 0 & (i > j) \end{cases}.$$

The Fisher information g appears in $D(P + dP, P)$ as the quadratic form

$$g = \sum_{k=1}^n \left(\frac{1}{\sigma_k} \sum_{i=1}^k r^{ik} d\mu_i \right)^2 + 2 \sum_{i=1}^n \left(\frac{d\sigma_i}{\sigma_i} \right)^2 + \sum_{l=1}^{n-1} \sum_{k=l+1}^n \left(\frac{\sigma_l}{\sigma_k} \sum_{i=l+1}^k r^{ik} dr_{li} \right)^2,$$

which is presented by a block diagonal $\text{diag}(g_{\mu\mu}, g_{\sigma\sigma}, g_{rr,2}, \dots, g_{rr,n})$, where

$$g_{\mu\mu} \left(= [g_{\mu_i, \mu_j}] \right) = \left[\sum_{k \geq i, j} \frac{r^{ik} r^{jk}}{\sigma_k^2} \right] = C^{-1} C^{-T}, \quad g_{\sigma\sigma} = \text{diag} \left(\left(\frac{2}{\sigma_i^2} \right) \right),$$

and $g_{rr,l} = [g_{r_{li}, r_{lj}}]_{i,j > l} = [\sigma_l^2 g_{\mu_i, \mu_j}]_{i,j > l}$ ($l = 1, \dots, n-1$). Lowering the upper indices of the α -connection by $\sum_L g_{KL} \Gamma_{IJ}^\alpha = \Gamma_{\{L,J\},K}^\alpha$, we have

$$\begin{aligned} \Gamma_{\{\mu_i, \mu_j\}, \sigma_k}^0 &= -\Gamma_{\{\mu_i, \sigma_k\}, \mu_j}^0 = \frac{r^{ik} r^{jk}}{\sigma_k^3}, \\ \Gamma_{\{\sigma_i, \sigma_i\}, \sigma_i}^0 &= \frac{-2}{\sigma_i^3}, \\ \Gamma_{\{\mu_i, \mu_j\}, r_{ab}}^0 &= -\Gamma_{\{\mu_i, r_{ab}\}, \mu_j}^0 = \sum_{k=b}^n \frac{r^{bk} (r^{ia} r^{jk} + r^{ik} r^{ja})}{2\sigma_k^2}, \\ \Gamma_{\{r_{li}, r_{lj}\}, \sigma_l}^0 &= -\Gamma_{\{r_{li}, \sigma_l\}, r_{lj}}^0 = \sum_{k \geq i, j} \frac{-\sigma_l r^{ik} r^{jk}}{\sigma_k^2}, \\ \Gamma_{\{r_{li}, r_{lj}\}, \sigma_k}^0 &= -\Gamma_{\{r_{li}, \sigma_k\}, r_{lj}}^0 = \frac{\sigma_l^2 r^{ik} r^{jk}}{\sigma_k^3} \quad (k \geq i, j), \\ \Gamma_{\{r_{li}, r_{lj}\}, r_{ab}}^0 &= -\Gamma_{\{r_{li}, r_{ab}\}, r_{lj}}^0 = \sigma_l^2 \Gamma_{\{\mu_i, \mu_j\}, r_{ab}}^0 \quad (a > l), \\ \Gamma_{\{L,J\},K}^0 &= 0 \quad (\text{for the other choices of } \{L, J\} \text{ and } K), \end{aligned}$$

and

$$\begin{aligned}
 \Gamma_{\{\mu_i, \sigma_k\}, \mu_j}^1 &= 2\Gamma_{\{\mu_i, \sigma_k\}, \mu_j}^0 \\
 \Gamma_{\{\sigma_i, \sigma_i\}, \sigma_i}^1 &= 3\Gamma_{\{\sigma_i, \sigma_i\}, \sigma_i}^0 \\
 \Gamma_{\{\mu_i, r_{ab}\}, \mu_j}^1 &= 2\Gamma_{\{\mu_i, r_{ab}\}, \mu_j}^0 \\
 \Gamma_{\{r_{li}, r_{lj}\}, \sigma_l}^1 &= 2\Gamma_{\{r_{li}, r_{lj}\}, \sigma_l}^0 \\
 \Gamma_{\{r_{li}, \sigma_k\}, r_{lj}}^1 &= 2\Gamma_{\{r_{li}, \sigma_k\}, r_{lj}}^0 \quad (k \geq i, j), \\
 \Gamma_{\{r_{li}, r_{ab}\}, r_{lj}}^1 &= 2\Gamma_{\{r_{li}, r_{ab}\}, r_{lj}}^0 \quad (a > l), \\
 \Gamma_{\{I, J\}, K}^1 &= 0 \quad (\text{for the other choices of } \{I, J\} \text{ and } K),
 \end{aligned}$$

and thus we also have

$$\begin{aligned}
 \Gamma_{\{\mu_i, \mu_j\}, \sigma_k}^{(-1)} &= 2\Gamma_{\{\mu_i, \mu_j\}, \sigma_k}^0 \\
 \Gamma_{\{\sigma_i, \sigma_i\}, \sigma_i}^{(-1)} &= -\Gamma_{\{\sigma_i, \sigma_i\}, \sigma_i}^0 \\
 \Gamma_{\{\mu_i, \mu_j\}, r_{ab}}^{(-1)} &= 2\Gamma_{\{\mu_i, \mu_j\}, r_{ab}}^0 \\
 \Gamma_{\{r_{li}, \sigma_l\}, r_{lj}}^{(-1)} &= 2\Gamma_{\{r_{li}, \sigma_l\}, r_{lj}}^0 \\
 \Gamma_{\{r_{li}, r_{lj}\}, \sigma_k}^{(-1)} &= 2\Gamma_{\{r_{li}, r_{lj}\}, \sigma_k}^0 \quad (k \geq i, j), \\
 \Gamma_{\{r_{li}, r_{lj}\}, r_{ab}}^{(-1)} &= 2\Gamma_{\{r_{li}, r_{lj}\}, r_{ab}}^0 \quad (a > l), \\
 \Gamma_{\{I, J\}, K}^{(-1)} &= 0 \quad (\text{for the other choices of } \{I, J\} \text{ and } K).
 \end{aligned}$$

The coefficients for the e-connection all vanish with respect to the natural parameter $\theta = (C^{-1}C^{-T}\mu, \xi)$, where $\xi = (\xi_{ab})_{1 \leq a \leq b \leq n}$ is the upper half of $C^{-1}C^{-T}$. Dually, the coefficients for the m-connection all vanish with respect to the expectation parameter $\eta = (\mu, \nu)$, where $\nu = (\nu_{ab})_{1 \leq a \leq b \leq n}$ is the upper half of $C^TC + \mu\mu^T$. Now we fix the third component r of (μ, σ, r) , and change the others. We take the natural projection $\pi : U = \mathbb{H}^n \times \mathbb{R}^{n(n-1)/2} \rightarrow \mathbb{R}^{n(n-1)/2}$ and modify the coordinates (μ, σ) on the fiber $L(r) = \pi^{-1}(r)$ into (m, s) in the next proposition.

Proposition 1. *The fiber $L(r) = \pi^{-1}(r)$ is an affine subspace of U with respect to the e-connection ∇^1 . It can be parametrized by affine parameters $\frac{m_i}{s_i^2}$ and $\frac{1}{s_i^2}$, where $m = [r^{ij}]^T \mu$ and $s = \sqrt{2}\sigma$.*

The fiber $L(r)$ satisfies the following two properties.

Proposition 2. *$L(r)$ is closed under the convolution $*$ and the normalized pointwise product \cdot between the probability densities.*

Proposition 3. *The fiber $L(r)$ with the induced metric from g admits a Kähler complex structure.*

We write the restriction $D|_{L(r)}$ of the premetric D using the coordinates (m, s) as

$$D|_L((m, s), (m', s')) = \frac{1}{2} \sum_{i=1}^n \left\{ \left(\frac{m'_i}{s'_i} - \frac{m_i}{s_i} \right)^2 + \frac{s_i^2}{s'^2_i} - 1 - \ln \frac{s_i^2}{s'^2_i} \right\}.$$

We take the product $U_1 \times U_2$ of two copies of the space U . Then the products $L_1(r) \times L_2(R)$ of the fibers foliate $U_1 \times U_2$. We call this the primary foliation of $U_1 \times U_2$. For each $(r, R) \in \mathbb{R}^{n(n-1)}$, we have the coordinate system (m, s, M, S) on the leaf $L_1(r) \times L_2(R)$. From the Kähler forms

$$\omega_1 = 2 \sum_{i=1}^n \frac{dm_i \wedge ds_i}{s_i^2} \quad \text{and} \quad \omega_2 = 2 \sum_{i=1}^n \frac{dM_i \wedge dS_i}{S_i^2}$$

respectively, on $L_1(r)$ and $L_2(R)$, we define the symplectic forms $\omega_1 \pm \omega_2$ on $L_1(r) \times L_2(R)$. We fix their primitive 1-forms

$$\lambda_{\pm} = 2 \sum_{i=1}^n \left(\frac{dm_i}{s_i} \pm \frac{dM_i}{S_i} \right).$$

The symplectic structures on the primary foliation defines a pair of regular Poisson structures. Now we take the $2n$ -dimensional submanifolds

$$F_{\varepsilon, \delta} = \left\{ \frac{m_i}{s_i} + \frac{M_i - \varepsilon_i}{S_i} = 0, s_i S_i = \delta_i \ (i = 1, \dots, n) \right\}$$

of the leaf $L_1(r) \times L_2(R)$ for $\varepsilon \in \mathbb{R}^n$ and $\delta \in (\mathbb{R}_{>0})^n$. The secondary foliation of $U_1 \times U_2$ foliates any leaf $U(r) \times U(r)$ by the $3n$ -dimensional submanifolds $F_{\varepsilon} = \bigcup_{\delta \in (\mathbb{R}_{>0})^n} F_{\varepsilon, \delta}$ for $\varepsilon \in \mathbb{R}^n$. The tertiary foliation of $U_1 \times U_2$ foliates all leaves F_{ε} of the secondary foliation by the $2n$ -dimensional submanifolds $F_{\varepsilon, \delta}$ for $\delta \in (\mathbb{R}_{>0})^n$. We take the hypersurface

$$N = \left\{ (m, s, M, S) \in L_1 \times L_2 \mid \prod_{i=1}^n (s_i S_i) = 1 \right\},$$

which inherits the contact forms $\alpha_{\pm} = \lambda_{\pm}|_N$. We can prove the following propositions.

Proposition 4. *With respect to the Kähler form $d\lambda_{-}$, the tertiary leaves $F_{\varepsilon, \delta}$ are Lagrangian correspondences.*

Proposition 5. *For any ε and δ with $\prod_{i=1}^n \delta_i = 1$, $F_{\varepsilon, \delta} \subset N$ is a disjoint union of n -dimensional submanifolds $\{s = \text{const}\} \subset F_{\varepsilon, \delta}$ which are integral submanifolds of the contact hyperplane distribution α_{+} on N .*

For each point $(\varepsilon, \delta) \in \mathbb{H}^n$, we have the diffeomorphism $\hat{F}_{\varepsilon, \delta} : \mathbb{H}^n \rightarrow \mathbb{H}^n$ sending $(m', s') \in \mathbb{H}^n$ to $(M, S) \in \mathbb{H}^n$ with $(m', s', M, S) \in F_{\varepsilon, \delta}$. We put

$$f_{\varepsilon, \delta}(m, s, M, S) = \frac{1}{2} \sum_{i=1}^n \left\{ \left(\frac{M_i - \varepsilon_i}{S_i} + e^{-h_i} \frac{m_i}{s_i} \right)^2 + e^{-2h_i} - 1 + 2h_i \right\},$$

where $h_i = -\ln \frac{s_i S_i}{\delta_i}$. Then we have

$$D|_L((m, s), (m', s')) = f_{\varepsilon, \delta}((m, s), \hat{F}_{\varepsilon, \delta}(m', s')).$$

For any $\zeta \in (\mathbb{R}_{>0})^{2n}$, we define the diffeomorphism

$$\varphi_{\varepsilon, \zeta} : (m, s, M, S) \mapsto (\zeta_{2i-1} m_i, (\zeta_{2i-1} s_i), (\varepsilon_i + \zeta_{2i} (M_i - \varepsilon_i)), (\zeta_{2i} S_i)),$$

which preserves the 1-forms λ_{\pm} . It is easy to prove

Proposition 6. *In the case where $\zeta_{2i-1} \zeta_{2i} = 1$ for $i = 1, \dots, n$, the diffeomorphism $\varphi_{\varepsilon, \zeta}$ preserves $f_{\varepsilon, \delta}$.*

For each $\varepsilon \in \mathbb{R}^n$, we take the set $f_\varepsilon = \{(f_{\varepsilon,\delta}, F_{\varepsilon,\delta}) \mid \delta \in (\mathbb{R}_{>0})^n\}$, and consider it as a structure of the secondary leaf F_ε . Then we can prove

Proposition 7. For any $\zeta \in (\mathbb{R}_{>0})^n$, the diffeomorphism $\varphi_{\varepsilon,\zeta}$ preserves the set f_ε for any $\varepsilon \in \mathbb{R}^n$. In the case where ζ satisfies $\prod_{i=1}^n (\zeta_{2i-1} \zeta_{2i}) = 1$, the diffeomorphism $\varphi_{\varepsilon,\zeta}$ also preserves the hypersurface N .

Hereafter we fix $\varepsilon = 0$. For any $\delta \in (\mathbb{R}_{>0})^n$, the diffeomorphism $\hat{F}_{0,\delta}$ interchanges the operation

$$(m, s) * (m', s') = \left(m + m', \left(\sqrt{s_i^2 + s_i'^2} \right) \right)$$

with the operation

$$(m, s) \cdot (m', s') = \left(\frac{m_i s_i'^2 + m_i' s_i^2}{s_i^2 + s_i'^2}, \frac{s_i s_i'}{\sqrt{s_i^2 + s_i'^2}} \right).$$

Namely,

Proposition 8. If $(m, s, M, S), (m', s', M', S') \in F_{0,\delta}$, then

$$\begin{cases} ((m, s) \cdot (m', s'), (M, S) * (M', S')) \in F_{0,\delta} \\ ((m, s) * (m', s'), (M, S) \cdot (M', S')) \in F_{0,\delta} \end{cases}.$$

A curve $(m(t), s(t)) \in \mathbb{H}^n$ is a geodesic with respect to the e-connection ∇^1 if and only if $\frac{m_i}{s_i^2}$ and $\frac{1}{s_i^2}$ are affine functions of t for $i = 1, \dots, n$.

Definition 1. We say that an e-geodesic $(m(t), s(t)) \in \mathbb{H}^n$ is intensive if it admits an affine parametrization such that $\frac{1}{s_i^2}$ are linear for $i = 1, \dots, n$.

Note that any e-geodesic is intensive in the case where $n = 1$. We show

Proposition 9. Given an intensive e-geodesic $(m(t), s(t)) \in \mathbb{H}^n$, we can parametrize its image

$$(M(t), S(t)) = \left(\left(\varepsilon_i - \frac{m_i(t) \delta_i}{s_i^2} \right), \left(\frac{\delta_i}{s_i} \right) \right)$$

under the diffeomorphism $\hat{F}_{\varepsilon,\delta}$ to obtain an intensive e-geodesic.

We have the hypersurface $N = \left\{ \prod_{i=1}^n s_i S_i = 1 \right\} \subset \mathbb{H}^n$ carrying the contact forms $\alpha_\pm = 2 \sum_{i=1}^n \left(\frac{dm_i}{s_i} \pm \frac{dM_i}{S_i} \right) \Big|_N$. We state the main result.

Theorem 1. The contact Hamiltonian vector field X of the restriction of the function $\sum_{i=1}^n \frac{m_i}{s_i}$ to the hypersurface N on any leaf $L_1(r) \times L_2(R) \approx \mathbb{H}^{2n}$ of the primary foliation of $U_1 \times U_2$ with respect to the contact form α_+ on N coincides with that for the other contact form α_- . The vector field X is tangent to the tertiary leaves $F_{\varepsilon,\delta}$ and defines flows on them. Here each flow line presents a correspondence between intensive e-geodesics as is

described in Proposition 9. Particularly, for $\varepsilon = 0$ and any $\delta \in (\mathbb{R}_{>0})^n$, the flow on the leaf $F_{0,\delta}$ presents the iteration of the operation $*$ on the first factor of $U \times U$ and that of the operation \cdot on the second factor.

Finally, we consider the transverse unitriangular group. We have the orthonormal frame

$$e_{ij} = \frac{\sigma_i}{\sigma_j} \sum_{k=j}^n r_{jk} \partial_{r_{ik}} \quad (1 \leq i < j \leq n)$$

with the relations $[e_{ij}, e_{kl}] = \delta_{il} e_{kj} - \delta_{kj} e_{il}$ of the unitriangular algebra. Using the dual coframe e^{ij} , the relations can be expressed as $de^{ij} = \sum_{k=i+1}^{j-1} e^{ik} \wedge e^{kj}$. The transverse section of the primary foliation of $U_1 \times U_2$ is the product of two copies of the unitriangular Lie group, which we would like to call the bi-unitriangular group. We fix the frame (resp. the coframe) of the transverse section consisting of the above e_{ij} (resp. e^{ij}) in the first factor U_1 and their copies E_{ij} (resp. E^{ij}) in the second factor U_2 . The quotient manifold carries the $(n-2)$ -plectic structure

$$\Omega = \sum_{i=1}^n e^{i,i+1} \wedge \dots \wedge e^{i,n} \wedge E^{n-i+1,n-i+2} \wedge \dots \wedge E^{n-i+1,n},$$

which satisfies $d\Omega = 0$ and $\Omega^n > 0$. We notice that, in the symplectic case where $n = 3$, the quotient manifold admits no Kähler structure (see [6]).

3. Discussion

It is remarkable that the transverse symplectic 6-manifold is naturally ignored in the Bayesian inference on 3-dimensional normal prior. The author conjectures that a similar geometry of 3 + 1-dimensional relativistic prior has some relation to the M-theory. See [7] for a relation between Poisson geometry and matrix theoretical and non-commutative geometrical physics.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Amari, S. *Information Geometry and Its Applications*; Springer: Tokyo, Japan, 2016.
2. Mori, A. Information geometry in a global setting. *Hiroshima Math. J.* **2018**, *48*, 291–306.
3. Hirzebruch, F. Hilbert modular surfaces. *Enseign. Math.* **1973**, *19*, 183–281.
4. Mori, A. A concurrence theorem for alpha-connections on the space of t-distributions and its application. *Hokkaido Math. J.* **2020**, in press.
5. Mori, A. Global geometry of Bayesian statistics. *Entropy* **2020**, *22*, 240.
6. Cordero, L.; Fernández, F.; Gray, A. Symplectic manifolds with no Kähler structure. *Topology* **1986**, *25*, 375–380.
7. Kuntner, N.; Steinacker, H. On Poisson geometries related to noncommutative emergent gravity. *J. Geom. Phys.* **2012**, *62*, 1760–1777.



© 2019 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).